

UNIVERSITY  
OF MICHIGAN

JAN 11 1962

SCIENCE  
LIBRARY

# Psi<sup>c</sup>ological Monographs

General and Applied

A Dual Approach to  
Rorschach Validation:  
A Methodological Study

By

James O. Palmer

Edited by Herbert S. Conrad

Published by The American Psychological Association

UNIVERSITY OF MICHIGAN LIBRARIES

# Psychological Monographs: General and Applied

*Editor*

**HERBERT S. CONRAD**

Federal Security Agency  
Office of Education  
Washington 25, D.C.

*Managing Editor*

**MARGARET K. HARLOW**

*Consulting Editors*

**DONALD E. BAIER**

**FRANK A. BEACH**

**ROBERT G. BERNREUTER**

**WILLIAM A. BROWNELL**

**HAROLD E. BURTT**

**JERRY W. CARTER, JR.**

**CLYDE H. COOMBS**

**ETHEL L. CORNELL**

**JOHN G. DARLEY**

**JOHN F. DASHIELL**

**EUGENIA HANFMANN**

**EDNA HEIDBREDER**

**HAROLD E. JONES**

**DONALD W. MACKINNON**

**LORRIN A. RIGGS**

**CARL R. ROGERS**

**SAUL ROSENZWEIG**

**KENNETH W. SPENCE**

**ROSS STAGNER**

**PERCIVAL M. SYMONDS**

**JOSEPH TIFFIN**

**LEDYARD R. TUCKER**

**JOSEPH ZUBIN**

MANUSCRIPTS should be sent to the Editor. For suggestions and directions regarding the preparation of manuscripts, consult the following article: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. Background and bibliographic materials must, in general, be totally excluded, or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as subscriptions and sales, change of address, author's fees, etc.) should be addressed to the American Psychological Association, 1515 Massachusetts Ave., N.W., Washington 5, D.C.

VOLUME 65  
NUMBER 8

WHOLE NO. 325  
1951

# Psychological Monographs:

## General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*  
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

---

## A Dual Approach to Rorschach Validation: A Methodological Study

By

JAMES O. PALMER

*School of Medicine, Washington University, St. Louis, Missouri*

*Accepted for publication September 21, 1950*

Price \$1.00

*Published by*

THE AMERICAN PSYCHOLOGICAL ASSOCIATION  
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

UNIVERSITY OF MICHIGAN LIBRARIES

COPYRIGHT, 1951, BY THE  
AMERICAN PSYCHOLOGICAL ASSOCIATION

## TABLE OF CONTENTS

I. THE TWO APPROACHES—A GENERAL STATEMENT .....	1
A. Introduction .....	1
B. The Interpretation as the Object of Validation .....	2
C. The Matching Approach .....	2
D. An Item Analysis Method .....	3
E. Purposes and Procedures of the Present Study .....	5
II. THE TEST, THE CRITERION, AND THE SAMPLE .....	6
A. The Selection of the Rorschach Technique .....	6
1. The Administration of the Test .....	6
2. The Protocols .....	6
B. The Subjects .....	6
C. The Criterion .....	7
III. THE MATCHING APPROACH .....	9
A. The Nature of the Interpretations .....	9
B. The Reliability of the Interpretations .....	9
C. The Reliability of the Therapists in Matching .....	10
D. Selection of the Matching Groups .....	10
E. The Sequence in Which the Two Approaches Were Used .....	11
F. How the Therapist Made His Selections .....	11
G. Results of the Matching .....	12
IV. THE CHECK LIST APPROACH .....	15
A. Selection of the Items .....	15
B. Arrangement of the Items on the List .....	18
C. The Number and Order of the Choices .....	18
D. The Instructions .....	18
E. Reliability of the Therapists in the Use of the Check List .....	18
F. Reliability of the Rorschach Judges in the Use of the Check List .....	18
G. Results of Validation on the Check List .....	21
V. THE RELATIONSHIP BETWEEN THE TWO APPROACHES .....	24
VI. SUMMARY .....	25
A. The Check List Approach .....	25
B. The Matching Approach .....	25
REFERENCES .....	27



## CHAPTER I

### THE TWO APPROACHES—A GENERAL STATEMENT OF THE PROBLEM<sup>1</sup>

#### A. INTRODUCTION

THE NUMEROUS investigations of the validity of the Rorschach have been reviewed very thoroughly in three articles by Hertz (5, 6, 7), and the validity of the TAT has been similarly summarized by Tomkins (19). However, a brief restatement of the methods used by various investigators in establishing the validity of projective techniques may serve as an orientation to the particular questions considered in the present study. Thus far, the authorities who have reviewed this problem have been concerned primarily with the types of evidence used for the validation of projective techniques. Hertz (5) distinguished four main types of validation studies, as providing different kinds of evidence: (a) *clinical studies*, in which the usefulness of these techniques is illustrated in the analysis of case histories, therapy, etc.; (b) *experimental studies*, in which changes in the test results are shown to accompany controlled changes in the individual's pattern of behavior; (c) *studies of defined groups*, in which certain patterns of test results are established as associated with the characteristic behavior of known groups; and (d) *predictive studies*, in which the degree of agreement is measured between the description of personality derived from the results of

a projective technique and that obtained from an analysis of some criterion, for example, a life history.

Although various writers (Tomkins [17], Macfarlane [12], and Symonds and Krugman [16]) have granted the possibility of approaching the validation of projective techniques in various ways, they have, at the same time, been careful to emphasize that the chosen method must take into account the nature of the technique being validated, particularly the concept of personality underlying the use of this technique. The most comprehensive argument concerning this point has been presented by Frank (3) in his classic discussion of the scientific basis of projective techniques. He pointed out that the "personality" which projective techniques are designed to evaluate is a *framework* of intervening concepts, a framework that relates the details of the individual's manifest behavior in terms of a *pattern* of motivations and attitudes. Macfarlane (12) also has considered this use of *interrelated* constructs to be a central problem "inherent in the validation of projective techniques." The point stressed by Frank is that personality as a configuration of functioning processes cannot be meaningfully broken up into isolated traits, but that part functions can only be described in terms of their interrelationships within the whole pattern. From this viewpoint, a description of personality derived from the results of a projective technique would require a method of validation which could test the accuracy of this description as a whole unit.

This assumption concerning the rela-

<sup>1</sup> This study was conducted under the auspices of the Veterans Administration Regional Office, San Francisco, while the author was in training in the Clinical Psychology Training Program of the Veterans Administration. The author is therefore deeply indebted to the Veterans Administration for making this study possible. The opinions stated in this report are, however, those of the author and do not necessarily reflect the viewpoint of the Veterans Administration.

tionship between personality descriptions derived from projective techniques, and the method employed for their validation constituted the point of departure of the present research. In the light of this concept of personality, two divergent predictive approaches to validation were applied to an established projective technique, the Rorschach. One of these approaches, the matching method, was designed specifically to test the validity of description of personality as whole units. The other approach, which attempts to validate these descriptions item by item, does not necessarily take into account the Gestalt nature of these descriptions. The general intent of this study was to test the relative applicability of these two methods in investigating the validity of a projective technique.

#### B. THE INTERPRETATION AS THE OBJECT OF VALIDATION

Before proceeding with the description of these two methods, it may be well to emphasize that this study is concerned with methods of validating the *interpretation* or description of personality as derived from the responses of the subject, rather than with consideration of specific *scores* or discrete responses. While there are merits in dealing with the so-called objective data of projective tests, this author agrees with such authorities as Hertz (6) and Macfarlane (12) that behind such scoring systems lie implicit assumptions about personality functioning. It thus appears to this author more reasonable to deal directly with these interpretative assumptions and avoid the current controversies concerning scoring categories and their discrete meanings. The question of whether or not a set of idiosyncratic responses represents in a rough manner the general pattern of functioning of the individual, and of

how this question may be answered, seems a legitimate object of study.

#### C. THE MATCHING APPROACH

It was clearly apparent to Vernon (19), even during the developmental stage of projective techniques, that there was a need for a statistical approach which would treat their interpretation as a single, whole unit. With this specific problem in mind, he developed what has become known in the literature as the *matching method*. As Vernon (18), Hunter (9), and Krugman (11) have used it, this method consists of the following procedures: An interpretative report from a projective technique and a case analysis are prepared, independently, for each individual in a given sample. The sample is divided into small groups, ranging from five to ten subjects, known as *matching groups*. The interpretative reports and case analyses of each group<sup>2</sup> are presented, unidentified as to subject, to several judges who then attempt to match each of the test reports to the corresponding case analysis of the same individual's life history. Validity is then expressed in terms of the success of this matching. Chapman (1) derived the statistics for determining the chance variation and the significance of the success of matching. Vernon (19) has added a formula for a coefficient of contingency, *C*, permitting a statement of the degree of relationship between the test and the criterion as implied in the success of the matching.

Vernon (18) admitted that his method was "only a coarse beginning" to the validation of projective techniques and suggested two additional steps to this

<sup>2</sup> While most investigators have used matching groups consisting of an equal number of reports and case analyses, the matching groups may be uneven, e.g., ten interpretations to five analyses, or five to one.

procedure: (a) the homogeneity of the matching groups must be determined (obviously, a group of very similar reports would be more difficult to match than a group of very dissimilar reports); and (b) the reliability of the matching judges should be determined.

The application of the matching method to the validation of projective techniques has produced varying results. In his original study on the Rorschach, Vernon (18) reported an average contingency coefficient,  $C$ , of  $.833 \pm .0315$ . Vernon noted that "the actual size of the  $C$  depends very largely on the degree of heterogeneity or distinctiveness of the subjects in each group. As far as possible, a normal degree of heterogeneity was aimed at" by randomly selecting the cases for the groups out of the whole sample (18, p., 213). However, Vernon's matching groups may have been more heterogeneous than he assumed, as is suggested by the results of later studies. Hunter (9), in a study of fifty school children, reported that only five Rorschach reports were matched correctly by all four judges, and that each judge, singly, was successful in only 30 to 40 per cent of the matchings. She concluded, therefore, that matching was "of doubtful value" . . . "Calculated to differentiate only extreme cases" (9). Although this investigator did not state her method of selecting the matching groups, she did remark that the personality sketches were all very similar.

The chief objection to the matching method, however, is that it permits, at best, only a very general statement about the accuracy of an interpretative report, namely, the statement that on the whole the report is similar to the personality pattern depicted by the criterion. While it is reassuring to know that an essentially meaningful interpretation may be drawn from a projective technique, it would be even more satisfactory to know how well the personality configuration is delineated in an interpretative report. If the interpretative report is, in general, similar to the case analysis, successful matching may occur, even though the accuracy of many of the statements within the report is dubious. In fact, Cronbach (2) considered that the match-

ing method depends too much on the presence or absence of small clues. A proponent of the matching method might argue that, if the same personality pattern is, in general, described in both the interpretative report and in the case analysis, then the statements about particular functions of the personality would be likely to be similar in both descriptions. Unfortunately, the matching method does not provide any tests of this argument.

#### D. AN ITEM ANALYSIS METHOD

In a study of the validity of the TAT, Harrison (4) introduced a procedure which simultaneously checked both the *degree* and *area* of the accuracy of his interpretations. His interpreters wrote out an "itemized analysis" of the test protocols, i.e., lists of separate interpretative statements, and the case analyses were prepared in a similar fashion. The judges then compared the two sets of statements, *item by item*, denoting each item of the interpretation as "right," "wrong," or "?". This index of validity was significantly higher for his sample of interpretations than for a random group of interpretations, or for a group of "mock" reports, matched randomly with the same criterion. Cronbach (2)<sup>3</sup> described a validation design for projective techniques which is quite comparable to that employed by Harrison. Cronbach's conclusions, which might also be said to apply to Harrison's method, were that his type of approach (a) yielded a statistically sound test of significance, (b) "identifies objectively the accurate and inaccurate aspects of the prediction," and (c) permitted "identification of the types of cases for whom prediction is relatively

<sup>3</sup> Since Cronbach's article was published after the present study was completed, the particular design which he introduced was not originally considered in this investigation.

accurate" (2, p. 373).

It should be noted that in this procedure, the proof of validity hinges on the premise that the judges accept the statements of the interpretation as being similar to the items in the case analysis. The criterion for being "similar" is not stated in either of these articles, and from this procedure, no conclusion can be drawn as to the *degree* of similarity between the two sets of items. This degree of similarity might be measured by a rating scale of agreement, as used by Krugman (11). The ultimate step in this validation procedure would be to demonstrate that the personality of the individual as inferred from the test and from a life situation could be described by the same statements, e.g., on a rating scale or on a check list of commonly used statements.

The feature of this "item analysis" method of greatest import to our discussion is that on the surface it contradicts the assumption by Frank (3) and Vernon (18, 19), namely, that, since these descriptions deal with an integrated personality structure, they could not be validated piece by piece. This seeming contradiction might be explained, however, by the hypothesis that *the validity of these separate items depends on the validity of the whole description*. In the strictest sense, this hypothesis would mean that separate statements within the interpretative report would be valid *only* if the whole interpretation were valid. At least, it would indicate that if the interpretation as a whole is accurate, then the isolated statements drawn from the interpretative reports would be *more likely* to be accurate. It should be emphasized that this hypothesis refers to the validation of interpretations which stress the *relationships* between the various functionings

within the personality as a whole.

However, the studies of Harrison (4) and by Cronbach (2) do not supply any direct evidence in support of this hypothesis, since neither study provided a test of the accuracy of the interpretations as integrated, whole reports. Harrison did not state whether or not certain of his cases had significantly lower "validity indices." Although Cronbach concluded that his procedure allowed "the identification of the types of cases for whom prediction is relatively accurate," he did not describe these cases in the part of his study reported to date. It is possible that in both of these studies, some interpretations were inaccurate in the description of the personality as a whole, and that, therefore, the items drawn from these interpretations failed to attain significance.

Unfortunately, it cannot be determined from either Harrison's or Cronbach's articles exactly on what basis the interpretations were broken down into isolated items. One main criticism of these two studies is that no rationale is presented as a basis for the selection of the items. In fact, Harrison did not adopt a dynamic, structural approach to personality in his interpretations, but stated that his approach to personality was "eclectic and emphasized common sense psychology" in contrast to Murray's (13) theories or to psychoanalysis. Nor did Cronbach describe the theoretical bias of his Rorschach interpreters, although he did indicate that a more complete report would follow his introductory article. Exactly what types of statements about personality were validated, or might be validated, in this fashion remains undetermined. There is no assurance that this methodological design is applicable to the validation of interpretations based

on a dynamic theory of personality.

#### E. PURPOSES AND PROCEDURES OF THE PRESENT STUDY

The purposes of the present study were:

1. To test the hypothesis (discussed above) that the validity of separate statements about personality, inferred from projective techniques, depends on the

accuracy of the interpretation as a whole.

2. To determine whether a test of the validity of isolated statements is applicable to interpretations based on a dynamic, structural concept of personality.

3. To determine whether the personality of the individual as inferred from the test protocol and from the criterion situation could be described by the same set of statements.

## CHAPTER II

### THE TEST, THE CRITERION, AND THE SAMPLE

#### A. THE SELECTION OF THE RORSCHACH TECHNIQUE

**I**N ORDER to study the applicability of two methods of validation, it was essential that these approaches be tested on a projective technique of relatively accepted validity. After due consideration, the Rorschach technique was chosen, mainly on the strength of a comparatively longer and more varied background of validation studies. For a complete review of these investigations of the validity of the Rorschach, the reader is referred to the three articles by Hertz (5, 6, 7).

Predictive studies of the accuracy of Rorschach interpretation have not been as numerous, nor have the results been as uniformly positive, as those reported for the other types of approaches. Other than the studies using the matching method—which is under consideration here—most statistical studies have attempted to correlate isolated Rorschach signs with manifest behavior, usually with negative results. In regard to these studies, Hertz remarked, "The abortive dissection of the psychogram in the search for static factors in isolation has distorted the [Rorschach] method" (6, p. 549).

In addition to summarizing the various studies containing evidence of the validity of the Rorschach, Hertz presented many positive criticisms of these studies and recommendations for further investigation. In particular, she stressed the need for more experimental and differential studies to evaluate the various hypotheses underlying the interpretation of the Rorschach. The main purpose

of her review was to stimulate sound experimental design in these studies.

#### 1. *The Administration of the Test*

The method of administration followed the procedure described by Klopfer and Kelley (10). All the examiners took particular pains to secure a thorough "inquiry" into the features of the blots which elicited the subjects' responses. Probing and suggestive questions were avoided, however, until the "testing of the limits." Twenty-one of the subjects were tested by the author; the other seven subjects had been previously tested by other experienced administrators.

#### 2. *The Protocols*

The validity of a Rorschach interpretation depends both on the quantity and quality of the subject's responses. Most of these records offer a wealth of variegated responses as raw data for the interpreter. The completeness of the records was, to some extent, assured by the technique of administration, i.e., by the thorough inquiry. The quality of the protocols may also have been affected by the nature of the sample; possibly the patients had been selected for psychotherapy because they were comparatively more responsive and less restricted in their functioning.

#### B. THE SUBJECTS

In three Veterans Administration installations, the Rorschach was administered to all patients who were currently receiving psychotherapy and to whom the Rorschach had not previously been given. Of these 28 subjects, 11 were in a

neuropsychiatric hospital, 11 were from an outpatient clinic, and 6 were attending a nearby university clinic. These subjects ranged in age from 19 to 42 years, with a mean age of 28 years. Ten of the patients were classified as psychotic, 16 as neurotic, and 2 had other diagnoses. Except for the differences in diagnostic classification (see further discussion below), the character of this sample did not appear to have any direct bearing on the study of these two methods of validation.

### C. THE CRITERION

Obviously, the validation of a projective technique depends on the use of a criterion description which is comparable in nature to the test interpretation, and which is based on an adequate study of the individual. While the functioning of the individual's personality may be inferred from his manifest behavior as summarized in a life history or factual case study, this functioning may be observed even more directly and intimately in a psychotherapeutic study, i.e., in the individual's expression of his feelings and attitudes during psychotherapy, and in his emotional reactions to the psychotherapeutic situation. This criterion has been used in at least two major clinical validations of projective techniques: Hertz and Rubenstein (8), and Tomkins (17). It was also recommended by Rosenzweig (15) in his outline for a comprehensive study of the validity of the Rorschach.

In the present study, the Rorschach interpretations were compared, in the two validation methods, with the therapists' impressions of their patients. The seventeen therapists acted as the judges in both validation experiments, i.e., they selected the Rorschach reports which matched their patients in the matching

experiment and described their patients in terms of the choices on the item check list. Thus, the validation judges were able to evaluate the Rorschach interpretations on the basis of an intimate and extensive knowledge of the subject, instead of having to reply on the basis of a summarized sketch compiled by a disinterested party.

The therapy which the patients were receiving was psychoanalytic in nature, i.e., its purpose was to reveal to the patient his unconscious attitudes and motivations through an analysis of his emotional reaction to the therapeutic situation itself. The purpose of this therapeutic study of the patient's underlying attitudes and motivations may be considered equivalent to the aim of Rorschach interpretation. In fact, these therapists frequently requested a Rorschach report on their patients' personalities as an aid in planning treatment (excepting the patients who were subjects of this experiment). A majority of the therapists were also expert in the administration and interpretation of the Rorschach technique.

In the main, the therapists' impressions of their patients were derived from frequent contact with them. The total number of therapeutic interviews at the time the therapist made his judgment ranged from 6 to 90, with a median of 19 interviews: only 4 subjects were interviewed less than 10 times, while 8 had been interviewed over 30 times. In 20 of the 28 cases, the Rorschach was administered when therapy was in a beginning stage, i.e., before the fifth interview: the median number of interviews at the time of testing was 1, with a range of 0 to 40. A median of 15 weeks elapsed between the time of testing and the time of judgment; in no case was there less than a 7-week interval, and in 2 cases, the in-

terval was over 30 weeks. During this period, the therapists interviewed their patients between 5 and 80 times, with a median of 11 interviews occurring between testing and judging. As to the frequency of these interviews, 6 of the cases were seen 3 or more times weekly, 5 others were seen twice weekly, and all but 2 patients were seen regularly at least once a week. These 2 cases were interviewed frequently, but at irregular intervals. The minimum opportunity which a therapist had to observe his

patient was 6 interviews, occurring at irregular intervals, over a period of 16 weeks. The maximum observation occurred in a case where the patient was interviewed 90 times, 3 times a week, over a 30-week period. In summary, it may be said that the therapists' impressions were derived after extensive and frequent contacts with the subjects and may be considered comparable, in their theoretical approach to personality, to the Rorschach interpretations.

## CHAPTER III

### THE MATCHING APPROACH

**I**N BRIEF, the matching experiment consisted of the following procedures:

1. Interpretative reports were prepared from each of the Rorschach records by one interpreter (the author).
2. The reliability of these reports was checked in two ways: by having the reports matched to the protocols, and by having them matched to a duplicate set of reports which had been prepared by another psychologist.
3. In order to test the reliability of the therapists in their matching technique, they were given a group of five sample interpretations from which they had to select the one which matched a corresponding sample case analysis.
4. For each patient, a matching group was chosen, consisting of the interpretation of that patient's Rorschach, referred to hereafter as the *experimental* interpretation; and of four other interpretative reports, to be referred to as *alternates*.
5. Each therapist was then asked to select, from the group of five reports, the one report which he believed most closely represented his patient; subsequently, the therapist made a second choice among the remaining four reports.

#### A. THE NATURE OF THE INTERPRETATIONS

In order to standardize the method of interpretation and the style of the interpretative reports, all the protocols were interpreted by one person, the author. These interpretations were based solely on an individual's responses to the test material and on his behavior during the administration of the test.<sup>1</sup> The method

<sup>1</sup> The patient's behavior during the testing (as distinguished from his responses to the materials)

of interpretation followed that outlined in Klopfer and Kelley (10), particularly in the scoring of the responses and in the preliminary analysis of the psychogram. The conceptual framework employed throughout this process of interpretation was, broadly speaking, psychoanalytical. As far as possible, these descriptions were couched in everyday idiom, and both Rorschach and psychoanalytic terms were avoided.

#### B. THE RELIABILITY OF THE INTERPRETATIONS

Since all of the interpretative reports employed in the matching experiment were prepared by one person, it was necessary to determine whether these interpretations were reliable in the same sense that the consistency and accuracy with which one scores a psychometric instrument might be checked. As a rough test of this reliability, three judges, skilled in Rorschach interpretation, attempted to match the reports to the protocols, in groups of five each. Since this study was directed at the reliability or consistency of fully verbalized interpretations rather than of standardized symbols or *scores*, no attempt was made to check the author's scoring. These judges were instead presented with the *unscored* responses and asked to match these directly to the author's statements about the various subjects' personalities. All three judges were 100 per cent successful in this matching.

Despite this positive result, it was pos-

was not recorded. Since most of the records were interpreted some time after the test administration, little if any account was taken of this factor.

sible to question the reliability of these interpretations, i.e., whether they were similar to descriptions derived by some other interpreter. As a further check on this reliability, the protocols were interpreted independently by another psychologist.<sup>2</sup> Three judges matched, with complete success, the first five interpretations of this second set of reports with the five corresponding interpretations by the author. Since this result coincided with the results of Krugman's (11) more comprehensive study of Rorschach reliability, the success of this single matching was considered sufficient indication of the reliability of the interpretations used in the present study.

### C. THE RELIABILITY OF THE THERAPISTS IN MATCHING

Prior to the matching of the Rorschach interpretations in the main part of this research, the therapists were briefly trained in the use of the matching technique. In order to check the reliability of the therapists in matching, the author prepared a case analysis on a patient not included in the validation study. This patient's Rorschach record was interpreted independently by the interpreter who had participated in the study of the reliability of the interpretations. Using this case analysis as a criterion, ten of the therapists attempted to select this experimental interpretation from among four alternative interpretations (previously prepared by this other interpreter).

On this trial, six of the therapists matched the sample interpretation correctly on their first choice, and three

<sup>2</sup> The author wishes to acknowledge the patient assistance of Mr. Mervin Freedman, Mr. Patrick Sullivan, and Mr. William Cook of The University of California who acted as the judges here, and of Mr. Allen Dittmann, who prepared the second set of reports.

others indicated it as their second choice. In this instance, successful matching on first choice could be expected by chance in two cases, i.e., two out of ten times. When first and second choices were considered, chance matchings might occur in four instances in ten matchings. According to the tables of "General Term of Poisson's Exponential Expansion" in Pearson (14)<sup>3</sup> the obtained results of both the first choices alone (six correct matchings) and of the two choices combined (nine correct matchings) are significant beyond the 5 per cent level of confidence. Admittedly, this limited study of the reliability of the therapists in matching a single case was not completely comparable to the matching study of the validity of the twenty-eight cases, as described below. Still, in view of the positive results of this brief reliability study, it seems reasonable to expect that these therapists would be approximately reliable in other matching experiments—particularly one in which they would be more familiar with the criterion, i.e., their own patients.

### D. SELECTION OF THE MATCHING GROUPS

As noted in Section I, the results of a matching study depends to a large extent on the variability among the descriptive reports which constitute the

<sup>3</sup> Throughout this study, many of the resultant proportions of chance agreement were so small that their distribution was thought to be considerably skewed and platykurtic. The use of a standard error of a proportion and its interpretation in terms of the normal probability integral would have yielded erroneous probabilities. It was thought, however, that the computation of exact binomial probabilities would have required more effort than their usefulness justified, so approximations to these probabilities were obtained from the Poisson distributions. This distribution is useful in approximately binomial probabilities when  $p$  is small in comparison to  $q$ , but where the possible number (of agreements, in our case) is finite.

matching groups. In contrast to previous investigations which also made use of the matching method, the present research included an attempt to control the heterogeneity of the matching groups. This particular step in the present research may, therefore, bear some detailed explanation.

The purpose of this step in the matching procedure was to select matching groups having the same degree of heterogeneity. It seemed desirable that the *alternate interpretations* to be included in a group with the experimental interpretation should be neither very different from nor very similar to that experimental interpretation. To achieve this degree of heterogeneity, it was necessary to compare each interpretation with all other interpretations which might appear within a group as an alternate. In order to estimate the differences between interpretations, the following rough rating scale was adopted:

*SS*—Both interpretative reports describe the same basic personality features and indicate many similar specific characteristics.

*S*—Both reports describe similar basic personality features but may differ in specific characteristics.

*SO*—Both reports describe some similar basic features and some similar specific characteristics, but also differ slightly in both respects.

*O*—Both reports differ as to basic features but present some similar specific characteristics.

*OO*—Both reports differ in all respects.

*X*—The reports are not comparable.

For the sake of convenience, the first fifteen reports collected from the sample were rated on this scale prior to the last thirteen reports. Three judges<sup>4</sup> rated these reports on the above scale; each judge made an independent rating first, followed by a final pooled rating by all three judges. These ratings were made on the over-all description of the personality rather than any specific cues. Thus, two interpretations discussing latent homosexual trends as important to the personality picture but differing in most other respects, i.e., in basic personality structures, might be rated at least *O*, if not *OO*.

In selecting the matching groups, reports which rated *SO* with the experimental report were given preference; a few reports compared as *O* or *S* were also used in some groups, but none of those extremely different or similar were included. Thus, the matching groups were

of appropriate heterogeneity with regard to the experimental report.

#### E. THE SEQUENCE IN WHICH THE TWO APPROACHES WERE USED

Since both the matching and the check list judgments were made by the same judges (i.e., the therapists), the sequence in which the two techniques were tested carried a possible contamination: a therapist who matched his case before using the check list might be influenced by the selected report when the time came for him to make choices on the check list. Or, if he used the check list first, he might acquire a set for the matching of the report. Although such bias could not be wholly prevented, its possible effect was taken into account by systematically varying the order in which the therapist performed the two tasks. Therefore, in fourteen of the cases (seven in each half of the sample), the therapist selected the report before he made choices on the check list—the sequence being reversed in the other fourteen cases. The possibility of this type of bias was further lessened by the fact that the therapists always made the two judgments separately, with an intervening period of two to three weeks.

#### F. HOW THE THERAPIST MADE HIS SELECTIONS

Each therapist was presented with one matching group for each of his patients. Each group consisted of five reports (one of which was derived from his own patient's Rorschach). The therapist was instructed to select the one report which matched his patient. After this first choice was made, the therapist was asked to name a second choice from the remaining four reports. The selection of a second choice was requested in order to

<sup>4</sup>The author appreciates the assistance of Mr. Timothy Leary and Mr. Walter Klopfer in this task.

allow for partial errors in matching, particularly in those instances when a judge might be undecided as to which of two similar reports to select.

Another factor which had to be considered in this procedure was the possibility that the patient's personality had been altered by the therapy which intervened between the time the Rorschach was administered and the time the therapist made his selection. Therefore, as he made his selection, the therapist was reminded of the date of the test administration, and he was asked to consider the patient's personality as it had been at that previous time.

#### G. RESULTS OF THE MATCHING

In eleven of the twenty-eight cases, the therapists correctly selected the interpretation of their patient's Rorschach as a first choice from the matching groups; only two more were correctly selected as second choices. In terms of chance expectancy (using Poisson's tables) this result is significant beyond the 3 per cent level of confidence. Using Vernon's (19) formula for the coefficient of contingency, *C* is equal to  $.434 \pm .078$ .

Although this matching was above chance, the relationship between the Rorschach and criterion, indicated by this *C*, was considerably lower than reported in previous studies. Vernon (18) found a *C* of  $.833 \pm .047$ .<sup>5</sup> Krugman (11) reported a *C* of  $.850$ . Both studies differ from the present investigation in two important aspects:

1. They did not specifically control the heterogeneity of their matching groups. It seems reasonable to presume that in the present study the control of this factor created a more difficult task for the judges, and consequently, a more acute test of the Rorschach reports.

2. The previous studies used equal numbers of reports and case analyses, while the present investigation employed a five-to-one matching. Thus, in the present study, the judges were forced to differentiate among five reports, with

<sup>5</sup> Vernon reported a *PE* of  $.0314$ , converted here for purposes of comparison to a standard error.

only one criterion as a basis of judgment; in this sense, the chance of success was probably much smaller than in the previous studies.

Considering the factor of a more differentiating task for the judges, who therefore had less chance for successful matching, the degree of validity obtained in the present study may perhaps be regarded as comparatively more significant than the findings in the previous studies which did not include these controls.

The results of the matching experiment might also have been affected by other variables in the nature of the sample or in any of the procedures used in collecting and presenting the data. Seven variables were considered as possibly affecting the results of this matching, namely: (a) the type of installation (hospital or outpatient); (b) the psychiatric diagnosis (psychotic or "other neuropsychiatric disorder"); (c) the total numbers of interviews (above or below the median of 19); (d) the numbers of interviews after testing (median or 11 interviews); (e) the frequency of the interviews (weekly or more frequent); (f) the order in which the Rorschachs were administered (a difference in results was indicated between the two halves of the sample); and (g) the judgment which the therapist made first (matching or check list). A study of the effect of these seven variables was made to discover if they had any possible relationship to the matching results (see Table 1).

Only one of these differences, the order in which the Rorschach tests were administered—and interpreted—is significant at less than the 1 per cent level of confidence. A possible explanation of this difference is that the reports in the last half of the sample might have been more incisive descriptions than the first fifteen interpretations. The ratings of heterogeneity, which were made in the procedure for selecting the matching groups, provided some measure of the qualitative differences among the reports—at least within each of the two halves of the sample, but, unfortunately, not over the entire sample. The results of this rating procedure, as shown in Table 2, indicate that in both halves of the sample, a significantly greater number of the comparisons were rated as different from one another (*O* or *OO*) than might be expected if the distribution of ratings had been even; on the other hand, the percentage of *S* and *SS* ratings was much less than expected. Thus, the efforts of the interpreter to achieve distinctive reports were sustained within each half of the samples. Whether or not this distinctiveness increased progressively from one half of the sample through the next cannot be stated conclusively inasmuch as not each interpretation was paired with every other interpretation throughout the whole sample. However, in view of the fact that no significant difference existed between the two halves

TABLE 1

DIFFERENCES, IN THE PROPORTION OF CASES MATCHED CORRECTLY, BETWEEN VARIOUS CHARACTERISTICS OF THE SAMPLE, AND BETWEEN VARIOUS PROCEDURES IN MATCHING

Groups Compared	N	Cases Matched Correctly		Diff.	CR	P
		No.	Prop.			
Hospital patients	11	4	.36	.05 ± .27	<1	
Outpatients	17	7	.41			
Psychotic patients	10	2	.20	.30 ± .14	2.15	.03
Other NP patients	18	9	.50			
Total Interviews:						
Over 19	14	5	.36	.07 ± .18	<1	
Under 19	14	6	.43			
Interviews after testing:						
Over 11	14	6	.43	.07 ± .18	<1	
Under 11	14	5	.36			
Frequency of interviews:						
Once weekly	15	6	.40	.02 ± .18	<1	
Over once weekly	13	5	.38			
Cases 1-15	15	2	.13	.56 ± .16	3.5	<.01
Cases 16-28	13	9	.69			
Check list first	14	7	.50	.14 ± .19	<1	
Matching first	14	5	.36			

of the sample in the proportion of *O+OO* ratings, it may be considered that the cases of the second half were no more distinctive, as compared among themselves, than those of the first half.

The second of these variables which showed a significant difference in the matching results was the diagnostic classification of the patients. Fewer of the cases diagnosed "psychotic" were matched correctly than those classed as "neurotic" or in other neuropsychiatric categories. Although the number of patients in these classifi-

cations was too small for further computation of differences, it was noted that seven of the ten psychotic cases fell in the first half of the sample. These results, if meaningful, would indicate that the interpretations of the Rorschach of the psychotic patients may have been less differentiating ones. In view of the fact that psychotic patients (other than paranoid types) often give vague and diffuse responses, it is to be expected that these interpretations would be less meaningful and distinctive. Such responses from the psychotic patients are consonant with

TABLE 2

DIFFERENCES BETWEEN OBTAINED AND EXPECTED RATINGS OF *S+SS* AND *O+OO*  
(Assuming an expected chance distribution of equal proportions of *SS+S*, *SO+X* and *OO+O*.)

Cases	Comparisons	Rating	Obtained	% Obtained	% Expected	Diff.	CR	P
1-15	105	<i>S+SS</i>	16	15.2	33±4.47	-17.8	3.73	<.01
		<i>O+OO</i>	44	48.8		+10.8	2.06	<.05
16-28	78	<i>S+SS</i>	9	11.5	33±4.9	-21.5	4.39	<.01
		<i>O+OO</i>	38	48.7		+15.7	3.21	<.01

the theory of personality used here, i.e., that inadequate perceptual differentiation is equated with psychoses. However, this concept is not helpful in distinguishing one psychotic patient from another, as was required of the matching judges. If the judges operated on the basis of this concept also, then the criteria may have been

as nondifferentiating as the Rorschach. Further study of the psychotic individual may be required, by both the Rorschach and other methods of observation, before a higher validity of interpretation can be demonstrated by the matching method.

## CHAPTER IV

### THE CHECK LIST APPROACH

THE CHECK LIST approach consisted of the following steps:

1. A list of thirty-four multiple choice items was constructed, consisting of statements commonly used in interpretative reports and in psychotherapeutic analysis.

2. The reliability of the therapists in the use of this check list was determined, using a sample case analysis as criterion.

3. Four Rorschach interpreters independently checked their choices on the multiple choice items for the twenty-eight Rorschach protocols. The reliability of these Rorschach judges was determined by computing the significance of the number of agreements between these judges, for each item.

4. The therapists checked choices on each item on the basis of their impression of their patients. The validity of the Rorschach judges' choices was then determined by computing the significance of the number of times that they agreed with the therapists on each item.

#### A. SELECTION OF THE ITEMS

In order to obtain a list of multiple choice items representative of the many abstractions used in Rorschach interpretation, each of the major categories of personality utilized in the interpretative reports was represented by at least one item. These six major areas of personality were as follows: (a) inner drives and attitudes, (b) emotional reactions and relationships, (c) sensitivity to emotional stimuli, (d) intellectual functioning and reality testing, (e) sexual attitudes and identification, and (f) anxiety and defenses against anxiety. Each of these

major categories or areas was further considered in four subdivisions or "dimensions": (a) the *frequency* or *extent* to which these areas were represented in his reaction to the test materials or to psychotherapy; (b) the characteristic *type* or *nature* of each area; (c) the *role* which each area played in the total pattern of the personality; (d) the *control* or manner in which the individual handled the attitude or reaction in question.

The questions asked in the interpretative reports concerning the individual's attitudes toward his identity and his inner motivations were represented on the check list by those items referring to fantasy life and inner drives, as follows:

(Quantity) No. 17. "Expression by the individual of his inner needs and drives, i.e., his striving for satisfaction of these drives, is: almost completely absent," to, "directly impulsive, showing an infantile lack of control."

(Role) No. 23. "Such inner fantasy life as the individual may allow himself is utilized for, or functions in his personality structure as: A. An internalization of certain unacceptable feelings, not permitted in overt behavior, e.g., for introduction of hostility in an intrapunitive manner. B. An attempt to organize and handle outer behavior in an integrated manner. C. A retreat from nearly all environmental frustrations, especially those in interpersonal relationships, with a handling of such relationships on a fantasy level. D. Very little, being poorly developed. E. Very little, being a source of anxiety in itself."

(Control) No. 19. "The method by which the individual handles and controls his inner emotional drives is chiefly: A. By fantasy solutions—possibly by divorcing such feelings from reality. B. By creative use of his energies, in a sublimated manner. C. By repressing them in a rigid and constricted manner. D. By direct release in overt behavior. E. By attempting to intellectualize, depersonalize or otherwise detach them from him emotionally."

Closely related to this general area of inner motivation are the individual's

attitudes toward his sexual functioning, which were sampled on the check list by the following items:

(Quantity) No. 6. "The extent to which the individual enters into heterosexual relationships is almost completely nil," to, "is so exaggerated as to pervade much of the individual's behavior."

(Type) No. 30. "The following attitude may be considered as the 'basic' one with which the individual regards his own sexual or 'sexualized' behavior: A. As an aggressive (sadistic act. B. As a dangerous (castrating) act. C. As a passive, receptive (incorporative) act. D. As a demonstration of potency, an egotistic self-assertion, (auto-erotic or exhibitionistic). E. As normal and acceptable (genital supremacy)." (A-D assume infantile sexual fixations or conflicts.)

No. 32. "The individual's general identification in most sexual and social roles is: A. with a dominant male figure. B. With a dominant female figure. C. With a passive male figure. D. With a passive female figure. E. Without a definite character and/or extremely ambivalent."

(Role) No. 10. "Homosexual relationships are utilized by the individual for, or play a role in his personality as: A. An integrated and mature part of his social behavior. B. A denial of rejection by other males. C. A denial of rejection by females. D. A minor role (e.g., for further satisfaction of narcissistic needs). E. An assertion of identification as to sexual role."

(Control) No. 24. "The chief method by which the individual handles his homosexual relationships is: A. By fairly overt emotional attachments possibly including sexual satisfactions. B. By repression of such feelings and/or avoidance of such relationships. C. By sublimating such feelings into socially acceptable channels of behavior. D. By retreating into fantasy solutions. E. By intellectually detaching the emotional aspects, depersonalization."

No. 34. "The individual handles, or reacts to, possible heterosexual relationships (or needs for such relationships) chiefly by: A. A retreat into fantasy (without necessarily breaking with reality). B. By accepting social restrictions, and sublimating where necessary. C. By repressing such drives, and/or depersonalizing them, detaching the emotional aspects of such relationships. D. By affective outburst—such as overt anxiety and panic, etc. E. By breaking with reality."

The manner in which the individual perceives and accepts the pressure of his environment was represented in the following items, which dealt with inter-

personal relationships and emotional reactions:

(Quantity) No. 27. "The degree to which the individual allows himself to become involved in emotional relationships with others is: a very limited involvement of any type," to, "a purely volatile and explosive reaction."

(Type) No. 25. "The emotional tone or affect which the individual displays in his emotional attachments with others is most often: A. Warm and spontaneous. B. (Absent.) C. Cold and detached. D. Hostile and oppositional. E. Forced and artificial.

(Role) No. 33. "Involvement in active emotional, interpersonal relationships serves the individual as, or plays a role in his personality structure as: A. A method of compensating for the inadequacies felt within himself. B. A release mechanism for the satisfaction of inner drives. C. A minor role—in an intraversive adjustment, under accumulated frustration or increasing environmental stimulation only. D. (None.) E. As a mature and integrated part of his behavior."

(Control) No. 19. "The principal method by which the individual handles and controls his emotional reactions in his interpersonal relationships is: A. By integrating them in a mature manner with other personal needs. B. By rigidly avoiding and denying the emotional aspects (isolation of emotion). C. By ignoring the reality of such an emotion and autistically withdrawing into fantasy solutions. D. By immature, and possibly aggressive, reactions. E. By depersonalizing such situations through intellectualizing or rationalizing."

Four other items were constructed requiring judgment relative to the individual's sensitivity to environmental stimulation:

(Quantity) No. 8. "The extent to which the individual allows himself to be receptive to the affective feelings of others, or to other emotional stimulation: A. Is limited and tenuous, chiefly when socially approved. B. Is practically absent. C. Is such that the individual is acutely aware of the emotional aspects of a situation. D. Shows a well-balanced and integrated sensitivity and tact. E. Shows a tendency to be unduly sensitive."

(Type) No. 9. "The individual's most characteristic reaction to the affective feelings of others and to the emotional stimulation from his environment, is: to be indifferent and disinterested," to, "to be overtly sensuous."

(Role) No. 16. "Sensitivity to the emotions of others or to other emotional stimuli is utilized by the individual for, or plays a role in his

personality as: A. (None.) B. A counter-reaction to repressed hostility. C. A withdrawal from frustration, from a more active emotional involvement. D. An integrated part of a mature handling of social relationships. E. A primary source of guilt and anxiety."

(Control) No. 2. "The way in which the individual controls and handles possible sensitivity to emotional stimulation, especially to the feelings of others, is usually: A. By rigid repression and restriction of such sensitivity. B. By attributing such stimulation to his own inner needs (by introjection). C. By reactively denying such feelings in aggressive, hostile behavior. D. By divorcing such feelings from reality and/or by fantasy solutions. E. By acceptance and integration of such sensitivity in social relationships."

From the area of intellectual functioning and of reality testing, the following items were derived:

(Quantity) No. 1. "The wealth of the individual's intellectual activity may be characterized as generally: impoverished, and tending to be perseverative," to, "having a wide range of interests, often being rich and original in content." No. 7. "The individual's intellectual productivity may be estimated as generally: very limited," to, "extensive."

(Type) No. 15. "The individual's intellectual approach to a problem or situation: A. Usually shows a tendency to abstract and over-generalize, without sufficient attention to everyday affairs. B. Tends to be overly critical, analytical, possibly picayunish. C. Usually shows a fair ability to conceptualize, but with adequate attention to practical concrete matters. D. May contain some evidence of delusional thought processes, forcing relationships between facts or distorting reality. E. Is most often a matter-of-fact approach, tending to be overly concrete."

No. 18. "The individual's ties to reality may be classified as chiefly: very strong, as never permitting any vagueness" through, "adequate—but not overly concerned with reality testing," through, "so tenuous as to easily become inadequate," to, "quite inadequate and/or absent."

(Role) No. 22. "Intellectual functioning is utilized by the individual for, or has a principal function in his personality structure as: A. A rigid defense against the release of inner drives and/or emotional ties with others, by depriving them of their emotional tone. B. A mature and normal mode of controlling himself and his environment. C. As an aid to autistic thinking, e.g., in delusional types of solutions. D. A highly aggressive, critical defense mechanism. E. Only a minor role, e.g., as an aid to immediate satisfaction of narcissistic needs."

(Control) No. 5. "The individual's contact with reality appears weakest: A. In his creative, inner fantasy life. B. In his active, and potentially aggressive, interpersonal relationships. C. In his sensitivity to emotional stimulation. D. In seemingly impersonal situations (to which affect has been displaced). E. In his release of instinctual drives."

The items included on the check list under the rubric of *anxiety* correspond in part to those considerations given ego-functioning above:

(Quantity) No. 28. "The degree to which the individual shows feelings of generalized disturbance may be estimated as: seldom more than a minimal and occasional uneasiness," to, "states of overwhelming panic."

(Type) No. 3. "The individual's expression of feelings of generalized disturbance may be characterized as: A. A free-floating type of anxiety state. B. A feeling of inner tension and conflict-guilt feelings. C. Overt depression. D. A sense of frustration and disappointment. E. (Relatively absent.)"

(Role) No. 21. "The effect of anxiety and/or guilt feelings on the individual's personality structure constitutes: no noticeable effect," to, "a gross breakdown of most functioning."

The different defenses were considered more specifically in the following items:

(Projection) No. 4. "Projection of guilt feelings onto others or the environment is used by the individual as a method of averting anxiety to the following degrees: very rarely," to, "extensively."

(Rationalization) No. 11. "The individual uses rationalization and justification as an intellectual evasion of anxiety to the following degree: very rarely," to, "extensively."

(Obsession) No. 12. "The individual uses compulsive behavior or obsessive thinking as a magical and ritualistic denial of anxiety to the following degree: very rarely," to, "extensively."

(Displacement) No. 14. "The individual attempts to avert anxiety by displacement of emotional content to some more 'neutral' situation: very rarely," to, "extensively."

(Withdrawal) No. 20. "The individual attempts to avert anxiety by fantasy solutions and/or by withdrawal from contact with reality to the following degree: very rarely," to "extensively."

(Normal reaction) No. 26. "The individual uses anxiety as a normal 'warning signal' of possible frustration: very rarely," to, "extensively."

(Acting out) No. 29. "The individual attempts to avert anxiety by 'acting out' of frustrations

onto the environment, by negativism and aggression, etc.: very rarely," to, "extensively."

(Isolation) No. 31. "The individual attempts to avert anxiety by rigid isolation of all emotional aspects of a situation: very rarely," to, "extensively."

#### B. ARRANGEMENT OF THE ITEMS ON THE LIST

It is possible that, if the items had been presented in the logical order of the scheme shown above, a judge's choice on one item might directly influence his choices on succeeding items, especially those within the same major category. In order to lessen the sequential effect, the items were presented in a random arrangement.

#### C. THE NUMBER AND ORDER OF THE CHOICES

For the sake of uniformity, five choices were listed for each statement. As was discovered afterward, this uniform number of choices was an unnecessary restriction; in many instances, a larger number of choices would have offered the judges more opportunity to describe their patients, and in a more accurate manner. In many instances, also, these five choices formed an obvious continuum, e.g., from "extensively" to "rarely," or "well adjusted" to "very disturbed," etc. Since a judge's choices on one item might well be influenced by the position on this continuum of his choices on previous items, the order of choices was varied in a random manner from item to item.

#### D. THE INSTRUCTIONS

The judges were instructed to make two choices on each item for each patient. Second choices were requested, because (a) it was possible that a pair of judges might agree, given two choices each, even though they might not agree on a first choice; and (b) some judges felt

less forced in their judgments when allowed two choices.

#### E. RELIABILITY OF THE THERAPISTS IN THE USE OF THE CHECK LIST

Reliability of the therapists in the use of the check list was determined in the same manner as was their reliability in the matching procedure. Using a sample case analysis (described above) as a criterion, ten of the therapists indicated choices on each item of the list. The therapists had previously had some training in the use of this type of list, in a trial run of a preliminary form. A rough estimate of the reliability of these therapists was obtained by assuming the agreement to be satisfactory when 5 or more therapists indicated the same first choice on an item—as describing this sample case history. When any two judges indicated the same two choices on any item, either as first or second choice, this was also noted as an agreement. For these "pairs" of choices, satisfactory agreement on an item was assumed when four or more judges used the same pair. On twenty of the items, five or more judges employed the same pair of choices. Although this reliability study was admittedly limited, it seems reasonable to assume that similar results might have been obtained if a more extensive study had been possible. Strictly speaking, the results of this step in the check list approach can only be taken to indicate that the therapists were able to agree satisfactorily on a majority of the items, using one sample case as a basis of judgment.

#### F. RELIABILITY OF THE RORSCHACH JUDGES IN THE USE OF THE CHECK LIST

The 28 Rorschach records of the sample were judged independently on the check list by four experienced Rorschach

interpreters. These judges<sup>1</sup> received a brief training in the use of this check list on a sample Rorschach protocol. The reliability of these Rorschach judges in the use of the check list was considered in terms of the number of agreements, i.e., the number of times they indicated the same choice on an item regarding the same individual. For each item, the 28 first choices of each judge were compared with those of every other judge—two judges at a time. Thus, six sets of comparisons were made for each item. The first choices of each pair of judges were tabulated on a five-square table, such that the agreements fell on the diagonal. The degree of agreement expected by chance was then computed on the assumption that the two sets of judgments were independent, subject to the restriction of the observed marginal totals. Since in most instances this chance degree of agreement was a relatively small proportion of the total number of cases—usually less than one-third—the significance of the difference between this number of agreements expected by chance and the number of agreements actually obtained was again read from the tables of Poisson's distribution (14). Where the expected number of agreements was larger than one-third of the total number of cases, the significance of this difference was computed in terms of the standard error of a proportion.

All in all, the four judges of the Rorschach agreed significantly about one-third of the time. Of the 204 comparisons for first choices (the choices of six pairs of judges compared on 34 items), 78 resulted in agreement at the 10 per cent

level or beyond; 51 of these were significant at the 5 per cent or beyond; and 26 at the 2 per cent level or better. For the two choices combined, 63 of the obtained agreements were significant at the 10 per cent level or beyond; 43 of these at the 5 per cent level; and 31 at the 2 per cent level or better. If we consider the rather abstruse wording of some of these items and the limitations on the number of choices, this degree of agreement is considerable. The large percentage of lack of agreement is not surprising in view of the fact that the judges were attempting to make almost *unqualified* statements about personality from such a restricted sampling of behavior, i.e., responses to ten ink blot pictures. The alternative or second choice did not appreciably increase the agreement between the judges.

Further analysis of the agreement and lack of agreement indicates that a fair number of items may be called reliable. If three significant agreements, i.e., agreement by three pairs of judges, be granted as indicating satisfactory reliability for an item, then about one-third of the 34 items may be called reliable—12 for first choices and 11 for the combined choices. An appreciable part of the lack of agreement may have been attributable to some particular pair of judges. However, it was found that all pairs of judges agreed to about the same degree. The degree of agreement within each area and within each dimension, as shown in Table 3, is summarized as a proportion of the total number of comparisons made in that area or dimension which showed significant agreement. The area in which the proportion of agreement was highest was intellectual functioning (.667). This particular area has been given close attention in the interpretative method of Klopfer and Kelley (10). Besides, it has also been thoroughly discussed in the current literature; therefore, the comparatively strong agreement shown here is not surprising. Two items which were included in this area, but which failed to have three or more agreements (Nos. 5 and 18) should properly belong to reality testing in general, rather than to specific intellectual functioning. Undoubtedly, a better definition of *reality testing* should have been reached by the judges.

<sup>1</sup> Mr. Walter Klopfer, Mr. J. Neil Campion, Jr., and Dr. Claire Thompson of the University of California were kind enough to devote many hours assisting the author in making these judgments.

TABLE 3  
DISTRIBUTION OF SIGNIFICANT AGREEMENTS BETWEEN RORSCHACH INTERPRETERS  
ACCORDING TO AREA AND DIMENSION OF PERSONALITY FUNCTIONING

Area	No. of Comparisons	Proportion of Agreement	Frequency		Type		Role		Control	
			Items	No. of Agreements	Items	No. of Agreements	Items	No. of Agreements	Items	No. of Agreements
Inner drives	18	.611	17	2	—	—	23	4	13	5
Sexual attitudes	36	.277	6	0	30	3	10	0	34	1
Emotional reactions	24	.417	27	4	25	2	33	1	1	3
Sensitivity	24	.458	8	3	9	1	16	1	2	6
Reality testing	36	.667	1	6	15	6	22	3	5	2
Anxiety reactions	18	.166	28	2	18	2	21	0	—	—
Number of comparisons			42		42		36		36	
Proportion of agreement			.524		.328		.250		.606	

The judges agreed reasonably well on another area: the one dealing with the individual's inner life, including his motivations and incorporated attitudes (.611). The significance of this agreement lies in the fact that this area is not as easily interpreted as that of intellectual functioning. This result might be explained by the hypothesis that the Rorschach is designed to tap these inner drives more than it does other areas. However, the degree of agreement about these inner drives does not appear to be much stronger than that about outer emotional reactions (.417) or sensitivity (.458). (The actual number of agreements was too small to permit statistical comparison.)

The lower percentage of agreement regarding emotional reactions and sensitivity may be traced to the disagreement among the judges about the dimensions of *type* and *role*. These two dimensions in these areas were the most difficult to restrict to five choices, and it seemed that greater agreement might have occurred if more choices had been provided.

Although the agreement on items concerning sexual attitudes (.277) might also have been improved by expanding the number of choices, the basic difficulty was more likely the confusion over the concepts employed in these items, particularly the concept "homosexual." This term was intended to refer to any type of attitude toward the same sex, rather than just to overt sexual behavior. Although the judges understood this broader connotation, they tended to limit their thinking about it to the latter, more common meaning.

The notably low proportion of agreement on the items pertaining to anxiety and defenses (.166) (see Table 4) is again explicable by the fact of the limited number of choices. To paraphrase: all anxiety is scarcely divisible into five parts! A more serious source of disagreement was found to be the poor definition of terms in this area. Even after the use of the check list had been reviewed among the judges, such terms as "displacement" were not consistently applied; note the significant disagreements on this item (No. 14) particularly. Several of the judges expressed the opinion that the items referring to defenses often overlapped in meaning or were otherwise confusing.

TABLE 4  
DISTRIBUTION OF SIGNIFICANT AGREEMENTS  
BETWEEN RORSCHACH INTERPRETERS,  
ACCORDING TO TYPE OF DEFENSE

Defense	Item No.	No. of agreements
Projection	4	2
Rationalization	11	0
Obsession	12	1
Displacement	14	-2
Withdrawal	20	0
Normal warning	26	2
Acting out	29	1
Isolation	31	2
Total No. of comparisons		48
Proportion of agreement		.166

Viewing the results according to *dimension*, the highest agreement occurred on those items concerning the individual's handling or *control* of his functioning (.606). Next on the scale of agreement came the dimension of *frequency* (.524). These results indicate that Rorschach interpreters can agree among themselves on the two following aspects: the way in which an individual controls his impulses, and the degree to which he uses any particular control or defense or expresses an attitude.

The difference in the percentage of agreement between *control* (.606) and *role* (.250) assumes an importance when one considers that both dimensions deal with *relationships* between reactions. Thus, the judges were in much better agreement as to the *control* relationship between various areas of personality than they were concerning the relative importance and relationship of a particular reaction or attitude in the individual's overall functioning. This lack of agreement about the concept of *role* may be due partially to inadequate phrasing of the items or limitations in the number of choices. But the fact should also be pointed out that Rorschach interpretation underscores this concept of *control*, and that, too often, little attention is given the importance of such control or defense in the "economics" of the personality.

Thus, if it be granted that Rorschach theory emphasizes the perceptual functioning of the ego, then the comparatively higher reliability on the items covering *intellectual functioning* and *control* was to be expected. Conversely, these results may be taken to show a lack of agreement on those aspects which are less clearly defined in Rorschach theory, namely, the

contentual factors, such as sexual attitudes and types of anxiety.

#### G. RESULTS OF VALIDATION ON THE CHECK LIST

The choices on each item by each of the four Rorschach judges were tabulated with those of the therapists—one pair of judges at a time; four differences between the obtained frequencies of agreement and that expected by chance were derived for first choices, and the significance of these differences was noted in terms of the approximation to binomial probabilities read from Poisson's tables.

Considering first choices alone, only 9 of the 136 comparisons (4 comparisons on the 34 items) resulted in significant differences, at the 10 per cent level of confidence or beyond. No significant agreement appeared for the combined choices (first and second choices together). Only 5 of the differences on first choices were "positive differences," i.e., the obtained agreement was larger than that expected by chance, while 4 were "negative differences," i.e., the obtained agreement being smaller than chance. Since, in a distribution of 136 differences, 13.6 might be expected to show such significance by chance, this small number of significant agreements cannot be considered to be of any statistical importance. The differences did not seem to occur in any meaningful pattern: no more than 1 difference occurred on any one item; almost an equal number of such differences occurred for each pair of judges; and these differences did not appear to have any relation to the scheme of personality areas and dimensions.

In general, these results support the hypothesis that the check list type of approach is not applicable to the vali-

dation of the interpretations of projective techniques when these techniques are interpreted by means of a dynamic concept of personality. Such statements as those used in this check list are apparently meaningless, except in the context of an integrated descriptive report.

Undoubtedly, the amount of agreement between the Rorschach judges and the therapists was dependent on the degree to which these two sets of judges agreed separately among themselves. The reliability studies discussed above showed that when one set of judges used one kind of data, either Rorschach records or therapy, the agreement in that instance was satisfactory. In terms of the number of items which showed significant agreement, the Rorschach judges agreed among themselves 35 per cent of the time, i.e., on 12 items, while the therapists agreed among themselves 58 per cent of the time, or on 20 items. In view of these reliabilities, there would seem to be some chance of obtaining higher validity. At the same time, one might inquire as to why these isolated statements were not applicable to the validation of Rorschach interpretation when they appear to be applicable to the study of its reliability.

Qualitative examination of the reliability of the Rorschach judges and of the therapists may serve to explain why there was agreement within each respective set of judges but little agreement between the two sets. In fact, the results of this reliability study may aid in the exploration of the relationships between the isolated statements and the whole reports.

As has been noted in the above discussion, the Rorschach judges agreed among themselves on those statements referring to the concepts which are most clearly

defined in Rorschach theory, namely, in the area of *intellectual functioning* and on the dimension of *control*. Although several other basic concepts are commonly recognized in Rorschach interpretation, these two concepts may be considered as the "axis" from which the whole pattern of the personality is evolved. It must be acknowledged that when other concepts are introduced in an interpretation, they are not based entirely on the relationship between those two primary inferences. However, the analysis of clues in the Rorschach protocol associated with other concepts than intellectual functioning or control is strongly influenced by the conclusions about these two concepts.

Although the therapists used the same general framework of concepts as adopted by the Rorschach interpreters, there was no direct evidence as to which particular concepts within this framework were central in each therapist's thinking. Since the therapy dealt principally with emotional relationships and with the analysis of emotional reactions, one might expect that the concepts concerned with this area of the personality determined the therapist's orientation. Thus, in considering a patient's personality as observed during therapy, the therapist might be inclined to give weight only to such intellectual functioning as was directly related to the patient's emotional life.

Although the Rorschach interpreter and the therapist may have had different concepts and clues in mind as they analyzed their respective observations of the patient's reactions, they were both attempting to infer a total picture of the individual's functioning, i.e., his personality structure. Each set of judges used their particular clues and concepts con-

sistently and reliably when considering their respective data. The two sets of judges did not agree significantly on the concepts which were not central to their separate considerations, especially when these concepts were isolated from the context of the whole structural pattern—as

was required on the check list. On the other hand, when the total picture of the individual was taken into consideration, agreement between the therapists and the Rorschach interpreter was obtained, as was demonstrated in the matching experiment.

## CHAPTER V

### THE RELATIONSHIP BETWEEN THE TWO APPROACHES

WE RETURN to the hypothesis which assumed that validity of (or agreements on) specific statements is more likely to be found when the whole descriptive report is validated. Bearing in mind that of the 28 whole reports in this study, only 11 were satisfactorily matched or validated, then, little or no validity could be expected throughout the check list *for the entire sample*. On the other hand, more agreement would be expected on the check list items for those 11 cases on whom there was also agreement on the whole reports. In order to test this hypothesis, the sample was divided into two groups, based on the results of the matching experiments, i.e., the 11 cases correctly matched, and the 17 cases in which matching was unsuccessful.

As one test of the hypothesis, the significance of the agreements between the Rorschach and criterion judges was calculated separately for these two groups on each item of the check list. This division yielded no significant agreement for any one of the items for either group, whether considering the first choices separately or the combined choices. These results were not surprising, considering the results of the validation of these items for the sample as a whole. In addition, a positive agreement significantly above chance was contraindicated because of the small number of cases in each of the two groups.

A preliminary analysis of the data indicated that the obtained agreements on the check list for the 11 matched cases

were consistently greater than those for the other 17 cases, although not enough to be statistically significant. This possible difference was again tested by contrasting the obtained agreements and disagreements for both groups on fourfold chi-square tables. Since the number of agreements on first choices was too small to permit this comparison for any item, a chi-square test was made on the first and second choices combined. No significant results were obtained on the whole. Only 11 of the 136 comparisons attained significance beyond the 10 per cent level of confidence. Five of these chi-squares indicated a difference in the expected directions, i.e., the correctly matched group had more agreements, but 6 were in the opposite direction, i.e., the group not correctly matched possessed greater agreement on the check list item in question than did the cases correctly matched. There was no indication that these results were associated with any particular pair of judges or with any definite area of content of the items.

The absence of any significant degree of difference between these two groups of cases indicates that the agreement between the Rorschach and therapy judges on the specific items had no relation to their agreement on the whole reports. The lack of significant agreement between the Rorschach judges and the therapists on the check list, for the whole sample, therefore, cannot be attributed to the presence of any group of cases in the sample which were not validly interpreted as whole reports.

## CHAPTER VI

### SUMMARY

#### A. THE CHECK LIST APPROACH

THE CHECK LIST approach was employed in this study for two purposes: (a) to study the possibilities of validating a projective technique on a set of isolated interpretative statements, and (b) to determine whether the behavior of the individual on the test and in a life situation could be described by the same statement. The results of validating an accepted projective technique, the Rorschach, on this check list indicate that this approach is not applicable to the validation of such projective techniques. In view of the fact that both the Rorschach interpreters and the therapists used this check list reliably when describing their respective observations, the main conclusion is that the check list approach may be applicable to the study of personality descriptions only when a common set of concepts is maintained as a reference point for inferring the total pattern of the individual's functioning.

In general, these findings support the contentions of Vernon (19), Frank (3), and other investigators who have argued that a description of the interrelated functioning of an individual can be validated only as a whole. In particular, it has been demonstrated that a total and integrated picture of the individual's personality may be valid, even though there may be no more than chance agreement between the judgments of a test interpreter and a criterion in regard to isolated statements about the individual's functioning.

This check list approach was a more rigid and exacting test of the validity of separate interpretative statements than

the "item analysis" design employed by Harrison (4) and by Cronbach (2). In the first place, neither of these investigators attempted to test whether the behavior of the individual in both the test situation and in the life situation could be described by a common set of statements. Secondly, in this check list approach, the isolated statements were selected according to a definite rationale, i.e., the scheme of "area" and "dimensions" of personality.

The results of the present study indicate that the behavior of the individual in both the test situation and a life situation could not be satisfactorily described by the same statement. Since the "item analysis" method does not make this demand on the test, it is probably a sounder approach. However, in the use of an "item analysis" approach it is recommended that the rationale for selecting statements from the whole reports be clearly stated.

#### B. THE MATCHING APPROACH

The chief advantage of the matching approach, according to Vernon (18), is that for whole interpretative reports, this approach tests the validity of the most essential features of the interpretation of projective techniques, i.e., the accuracy with which the *interrelated pattern* of the individual's functioning is described. The findings of the present study support Vernon's contention. By means of the matching approach, validity was demonstrated for descriptions of personality which emphasized these interrelationships. On the other hand, no validity was obtained when isolated interpretative

statements were applied in which the relationships between various functionings of the personality were not elaborated.

One of the major concerns in the present study of the matching approach was the effect of the heterogeneity of the matching groups. This study indicated that use of matching groups of an "optimum" heterogeneity resulted in a smaller number of successful matchings than were obtained in previous studies using randomly selected groups. Since the matching approach is essentially a test of *interindividual* differentiation, careful attention must be paid to the nature of the sample of individuals from whom

a particular individual is to be differentiated.

The chief criticism directed against the matching approach has been that it does not provide a test of the accuracy with which various part-functionings within this whole pattern are delineated. The present study attempted, without success, to test this *intraindividual* differentiation of an interpretation, by means of the check list approach. Since the completion of the present study, an "item analysis" design has been suggested by Cronbach (2) which, if used in conjunction with a matching approach, may provide a thorough statistical method for the validation of projective techniques.

## REFERENCES

1. CHAPMAN, D. W. The statistics of correct matching. *Amer. J. Psychol.*, 1934, **46**, 287-298.
2. CRONBACH, L. J. A validation design for qualitative studies of personality. *J. consult. Psychol.*, 1948, **12**, 365-375.
3. FRANK, L. K. Projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 398-413.
4. HARRISON, R. Studies in the use and validity of the TAT with mentally disordered patients. II. A quantitative validity study by the method of blind analysis. *Character & Pers.*, 1940, **9**, 122-138.
5. HERTZ, MARGUERITE R. The validity of the Rorschach method. *Amer. J. Orthopsychiat.*, 1941, **11**, 512-520.
6. HERTZ, MARGUERITE R. Rorschach: twenty years after. *Psychol. Bull.*, 1942, **39**, 529-572.
7. HERTZ, MARGUERITE R. The Rorschach method: Science or mystery? *J. consult. Psychol.*, 1943, **7**, 67-79.
8. HERTZ MARGUERITE R., AND RUBENSTEIN, B. A comparison of three blind Rorschach analyses. *Amer. J. Orthopsychiat.*, 1939, **9**, 295-315.
9. HUNTER, M. E. The practical value of the Rorschach test in a psychological clinic. *Amer. J. Orthopsychiat.*, 1939, **9**, 278-294.
10. KLOPFER, B., AND KELLEY, D. M. *The Rorschach technique*. New York: World Book Co., 1942.
11. KRUGMAN, JUDITH I. A clinical validation of the Rorschach with problem children. *Rorschach Res. Exch.*, 1942, **5**, 61-70.
12. MACFARLANE, JEAN W. Problems of validation inherent in projective methods. *Amer. J. Orthopsychiat.*, 1942, **12**, 405-410.
13. MURRAY, H. A., AND ASSOCIATES. *Explorations in personality: A clinical and experimental study of fifty men of college age*. New York: Oxford Univ. Press, 1938.
14. PEARSON, K. *Tables for statisticians and biometricalians*. Cambridge, Eng.: Cambridge Univ. Press, 1914.
15. ROSENZWEIG, S. An outline of a cooperative project for validating the Rorschach test. *Amer. J. Orthopsychiat.*, 1935, **5**, 395-403.
16. SYMONDS, P. M., & KRUGMAN, M. Projective methods in the study of personality. *Rev. educ. Res.*, 1944, **14**, 81-93.
17. TOMKINS, S. S. *The Thematic Apperception Test*. New York: Grune and Stratton, 1947.
18. VERNON, P. E. The significance of the Rorschach test. *Brit. J. med. Psychol.*, 1935, **15**, 199-217.
19. VERNON, P. E. Matching methods as applied to the investigation of personality. *Psychol. Bull.*, 1936, **33**, 149-177.